

Linguistische Datenverarbeitung

Das IDS benutzt, um seine sprachwissenschaftlichen Aufgaben zu erfüllen, in intensivem Maße die Datenverarbeitung mit dem Computer. Im Mittelpunkt steht das Anlegen, Verwalten und Erschließen großer Sammlungen von Texten, seien sie geschrieben oder gesprochen.

Die Datenverarbeitung legt nun solche Sammlungen von Texten im Computer an und macht sie nach den Regeln der Datenverarbeitungskunst befragbar. Dabei sind folgende Schwerpunkte gesetzt:

- COSMAS: Im Projekt CO(rpus) S(torage) M(aintenance) and A(ccess) S(ystem) werden geschriebene Texte gesammelt und als Volltextdatenbank organisiert. Sie sind nach Wörtern und beliebigen Wortkombinationen abfragbar. Der Benutzer erhält die Treffer, die zugehörigen Textstellen und die Quellennachweise. Umgesetzt auf DIN A4 stehen derzeit ca. 100.000 Seiten Texte on-line im Dialog für Recherchen zur Verfügung; bis 1994 soll diese Textmenge auf 500.000 Seiten anwachsen. Bereits jetzt bietet COSMAS eine große Textvielfalt: Goethe, Thomas Mann, Karl Marx, Grass, Die Zeit, Mannheimer Morgen, Neues Deutschland, Die Welt, Pörtlner »Die Erben Roms« und vieles andere mehr. Darun-

ter sind auch so zeitnahe Textsammlungen wie das »Wendekorpus«, eine Sammlung von Flugblättern und Zeitungsartikeln aus der Zeit um die Wiedervereinigung aus der ehemaligen DDR und aus der Bundesrepublik.

Ab Herbst dieses Jahres wird das COSMAS-System über öffentliche Datennetze auch für externe Nutzer zugänglich sein.

- IDOSSA: I(interaktives) Do(kumen)tations) S(ystem für) S(pracharchive) ist ein Verbundprojekt, das das IDS zusammen mit dem Süddeutschen Rundfunk und weiteren Partnern durchführt. Hier wird Entwicklungsarbeit geleistet, die es am Ende ermöglicht, gesprochene Sprache als Schallereignis digital im Computer abzuspeichern und als Datenbank verfügbar zu machen.

Das Programmsystem wird angewendet werden, um die Tonbänder der Spracharchive des IDS in ähnlicher Weise zugänglich zu machen wie die geschriebenen Textsammlungen (vgl. COSMAS), aber auch z. B., um die Erschließung der Originaltonarchive des Süddeutschen Rundfunks zu erleichtern und zu verbessern. Das Dokumentationsverfahren wird dann einsetzbar für Verfahren zur Automatisierung von Sendeabläufen.

- DIDA: Das Arbeitsvorhaben DI(skurs) DA(tenverarbeitung) speichert Gesprächsmitschriften und deren Bearbeitung und Kommentierung zeitsynchronisiert ab. Das Programmsystem steht in Kürze auch für Recherchen über öffentliche Datennetze zur Verfügung.

- GRADA: Es wird fortlaufend eine GR(ammatische) DA(tenbank) eingerichtet. Große Teile der Textsammlungen, über die oben berichtet wurde, sind durch ein Computerprogramm grammatisch analysiert worden. So kann man recherchieren, wo der Konjunktiv benutzt wird, welche Genitivattributionen üblich sind und vieles mehr. Ziel der Linguistischen Datenverarbeitung ist, den Computer als Hilfsmittel in die sprachwissenschaftlichen Arbeitsverfahren zu integrieren und nutzbar zu machen. Sie sieht sich in der Pflicht, die über 20jährige Tradition des IDS beim Einsatz des Computers in der Sprachforschung fortzuführen und auch zur Weiterentwicklung des nationalen und internationalen Standards in dieser Hilfsdisziplin der Sprachwissenschaft beizutragen.

Arbeitsstellenleiter: Robert Neumann;
wiss. Mitarbeiter: Dr. Cyril Belica, Dr. Martin Holena, Dr. Irmtraud Jüttner, Anton Schlatter, Dr. Rudolf Schmidt, Doris al-Wadi.